



CDD: Conserved Domain Database

Functional units and building blocks for protein families

<https://www.ncbi.nlm.nih.gov/cdd/>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope

The NCBI Conserved Domain Database (CDD) is a protein domain annotation resource, and includes a collection of ancient protein domains (> -0.5 Billion years old) and hierarchies of domains related by common descent. Each conserved domain (CD) record consists of a protein multiple sequence alignment, a consensus sequence, and a PSSM (position-specific score matrix) that quantitatively represents the information in the alignment. Whenever possible, 3D structural information is used to define and refine the alignment models. CDD is supported by an active curation effort that identifies new domains, builds domain hierarchies, and continually updates existing records with new sequences. In addition to NCBI curated entries, it also contains entries from Pfam, Smart, Protein Clusters, TIGR, and last release of COG.



Access to CDD data

CDD is part of the NCBI Entrez system and is extensively linked with other NCBI data. CDs can be found by direct text searching from the CDD homepage. Relevant CDDs present in a protein entry will be linked from that record. PSSMs of each CD are represented in scoremat format, and these scoremats have been made available for search with a protein query sequence through a specialized BLAST program called reverse psi-blast (RPS-BLAST). These scoremats are also available for download from the ftp site so that a selected subset can be used to build custom domain databases for local RPS-BLAST searches using the standalone BLAST+ package.

Direct text query of the CDD database:

<https://www.ncbi.nlm.nih.gov/cdd/>

Indirect query with a sequence using RPS-BLAST:

<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

RPS-BLAST access from BLAST Homepage:

<https://blast.ncbi.nlm.nih.gov/>

Download of scoremat data files by FTP:

<ftp://ftp.ncbi.nlm.nih.gov/pub/mmdb/cdd/>

Searching with Text Query

Entering a set of query terms in the search box and pressing the “Search” button (A) from the CDD homepage retrieves relevant records from the database. The “Advanced” page (B) provides access to indexing fields and their terms, which can be used to construct complex query terms for more specific retrieval.

Searching with (corA) AND “curated”[filter] retrieves group of conserved domain records shown in summary format (C). These domains are present in well-conserved family of genes commonly found in bacteria and archaea, that encode a bi-valent cation transporter for Mg²⁺ influx or Zn²⁺ efflux. The thumbnail (D) indicates that the domain contains a structure record and can provide additional information on sequence-structure relationship. Clicking the title (E) opens the full record.

Elements of a CDD Record

A CDD record contains several elements that provide key information for the domain.

- **Sequence alignment:** Each aligned sequence is matched to a record in Entrez Protein database. If sequences with structural data (from the PDB database) exist, one of them will be chosen as the master sequence of the alignment.
- **Consensus sequence:** Each position in the consensus sequence contains the residue with the highest frequency in that column of the alignment. For a column to be included in the consensus sequence, at least 50% of the sequences must have an aligned residue in that column.
- **PSSM:** For each position in the consensus, frequency ratios (expected/observed) are calculated for each amino acid, and these frequencies are converted to scores in a PSSM. The PSSM thus has the same number of columns as the consensus and precisely defines the extent of the domain.

A Sample CDD Record

In the detailed record view of a conserved domain record (shown at the right), the summary on the domain function and taxonomic distribution (A) is displayed at the top. Clicking the thumbnail (B) opens the protein structure in Cn3D to allow interactive examination of the sequence alignment in the context of the structure and its feature annotation. Portlets in the left column provide statistics and relevant links to other tools, such as PSSM Viewer and CD Tree. An embedded display of CD Tree is also shown to the right (C). The structure guided functional annotation of this domain is summarized in the "Conserved Features/Sites" section (D). Features and sites annotated are listed in individual tabs. The protein sequence alignment is shown at the end (E), where the color of the residuals indicates the level of conservation and "#" and yellow column background indicate a conserved feature selected (F).

cd12828: TmCorA-like_1

Thermotoga maritima CorA-like subfamily

This subfamily belongs to the Thermotoga maritima CorA (TmCorA) family of the MIT superfamily of essential membrane proteins involved in transporting divalent cations (uptake or efflux) across membranes. Members of this subfamily are found in all three kingdoms of life.

<https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=213362>

Links

- Source: cd12822
- Taxonomy: cellular organisms
- PubMed: 23 links
- Protein: Representatives, Specific Protein, Related Protein, Related Structure, Architectures
- Superfamily: d00459
- BioSystems: 28 links

Statistics

- PSSM-Id: 213362
- View PSSM: cd12828
- Aligned: 203 rows
- ThresholdBitScore: 281.626
- ThresholdSettingGi: 343953375
- Created: 17-Jul-2012
- Updated: 17-Jan-2013

Structure

Structure View

Program: Cn3D

Drawing: All Atoms

Aligned Rows: up to 10

Download Cn3D

Hierarchy

Interactive Display

Display: cd12828 Branch

Download CDTree

Conserved Features/Sites

oligomer | metal binding | metal binding | CI binding site | periplasmic

Feature 1: oligomer interface [polypeptide binding site]

Evidence:

- Structure: 2HN2: Thermotoga maritima CorA forms a homopentamer in which each monomer contacts two other monomers, contacts at 4A - View structure with Cn3D
- Citation: PMID 16902408

Download Cn3D for Viewing 3D Structure | Scroll to Sequence Alignment Display

cd12828 is part of a hierarchy of related CD models. Use the graphical representation to navigate this hierarchy. cd12828 is a member of the superfamily d00459.

cd12828 Sequence Cluster

Zoom In

Sub-family Hierarchy

Interactive Display with CDTree

- cd11744 MIT_CorA-like
- cd12821 EcCorA_ZntB-like
- cd12823 Mrs2_Mfn1p-like
- cd12824 ZntB-like
- cd12833 ZntB-like_1
- cd12834 ZntB_u1
- cd12825 EcCorA-like
- cd12835 EcCorA-like_1
- cd12836 HpCorA-like
- cd12837 EcCorA-like_u1
- cd12826 EcCorA_ZntB-like_u1
- cd12827 EcCorA_ZntB-like_u2
- cd12822 TmCorA-like
- cd12828 TmCorA-like_1**
- cd12829 Alr1p-like
- cd12830 MtCorA-like

Sequence Alignment

Reformat | Format: Hypertext | Row Display: up to 10 | Color Bits: 2.0 bit | Type Selection: the most diverse members

Feature 1

2HN2_A 61 TPTWINITGIhrtDVVQRVGEFFgIHPVLEDIlnvthQRPKVEFFenYVFIVLkmftydkn---lheleseQVSLILtk 136

gi 257170536 26 GTTWIHAPEvt-daELDAIAETFaIHPVLEDIlnvthQRPKVEFFenYVFIVLkmftydkn---lheleseQVSLILtk 103

gi 353194386 27 GTTWVRVTDps-eAELDEVADGYdLHPIEDVlg-dVRPKVEFFdgHTFVLLKaatlrngdvafadelrdqQVAFFLgd 104

gi 307611750 61 HNIWIDISGLadIEKITKICSEynIHPVLEDIlntrQRPKLEIVddYIFVVKllqspnn---hlsynteqFCLIIkq 136

gi 307156401 65 GSKWVIITGLantQLINDICEFYgMHPVLEDIlnthQRPKIESFddYIYIVIKvilpmed---egiysqASIIIfk 139

gi 257797392 56 WNHWIQICIGVhdpAVLEGFGRKHFgIHPVLEDIlntrQRPKLEHFdpSLFLVCKwlgfnd---rslsdnHISLIMqn 130

gi 88187722 66 GICWVHISGLkdiDKINQILSDAILHPVLEDIlntrQRPKLEHFdpSLFLVCKwlgfnd---rslsdnHISLIMqn 130

gi 110620367 27 YNRWISLTPe-kQDIDLVSFKfSLHPVLEDIlntrQRPKLEHFdpSLFLVCKwlgfnd---rslsdnHISLIMqn 99

gi 2649821 63 KKLWIDVVGvhdSLIAKICEFLgIHPVLEDIlntrQRPKLEHFdpSLFLVCKwlgfnd---rslsdnHISLIMqn 135

gi 282157735 27 CNKWVSMSPs-pQEFKAVADAFELHPVLEDIlntrQRPKLEHFdpSLFLVCKwlgfnd---rslsdnHISLIMqn 98

A Sample CDD Record (cont.)

Furthermore, the structure thumbnail (A) in the Feature section links to an annotated Cn3D record with residues involved in oligomerization highlighted (B), where yellow residues (in both structure and sequence panels) are involved in interaction with the same set of residues from adjacent units (in pink) in homopentamer formation. The annotation panel also provides access to other annotated features (C). Residues for the selected features can be marked by clicking the "Highlight" button (D).

Conserved Features/Sites | **PubMed References**

oligomer | **metal binding** | **metal binding** | **CI binding site** | **periplasmic**

Feature 1: oligomer interface [polypeptide binding site]
Evidence:
 ■ **Structure:** 2HN2: Thermotoga maritime CorA forms a homopentamer in which each monomer contacts two other monomers, contacts at 4A
 - View structure with Cn3D
 ■ **Citation:** PMID 16902408
[Download Cn3D](#)

CDD Annotations

Annotations: oligomer interface, metal binding site, metal binding site, CI binding site, periplasmic entrance, putative hydrophobic gate, basic sphincter

CDD Descriptive Items

Name: TmCorA-like_1
 TmCorA-like_1: This subfamily belongs to the Thermotoga maritima CorA (TmCorA)-family of the MIT superfamily of essential membrane proteins involved in transporting divalent cations (uptake or efflux) across membranes. Members of this subfamily are found in all three kingdoms of life. It is functionally diverse subfamily, in addition to the CorA Co2+ transporter from the hyperthermophilic Thermotoga maritima, it includes Methanosarcina mazei CorA which

TmCorA-like - Sequence/Alignment Viewer

View Edit Mouse Mode Unaligned Justification Imports

2HN2_A ~ i i r k k R A D Y L L S L I D A L V D D Y F V L L E K I D D E I D V L E E V I e r p ~
 gi 159893589 ~ r i l e r G A D ~ L A H A V L D A V V D E F L P Y L D R I D E E V D F L D S I i q d p ~
 gi 31790359 ~ i a c q r G A A F L L H Q I L D Q I V D Y S P V L D D F D E R I D Q L E A D I f l k t ~
 gi 332698498 ~ a l f e k G T D F I V Y N L M D A L V D E Y F L L D D I D E R L N R I G R I f k k p ~
 gi 312111851 ~ d v q q G P F H V M Y R I D K I V D D Y F P P I Y H I E D V L N E L E N A n d e t ~
 gi 344033684 ~ n i f e r G I D F L L Y N L A D E I I D D Y F P I V D E V G N K I D E A E I f e k e ~
 gi 302201007 ~ a v f q k G V D F L L Y S I S D T L V D E Y F L L D D L D E R L S I I E D S I f g r p ~
 gi 156234341 ~ h d f g e D V A E L L Y H L L D T I V D E Y F P I L D A L I E R V E D V E A Q I f d h f ~
 gi 269790061 ~ s k l k p S A G V L L Y S I I D S I V D E Y F P L V D T L G D H I E R F E E T I f s g r ~
 gi 319995057 ~ r l i t r G S D F L C H A I L D H L V D D Y F P L V D E I D E I Q F I E L D Q V I q s p ~
 gi 390174485 ~ i i r k k R A D Y L L S L I D A L V D D Y F V L L E K I D D E I D V L E E V I e r p ~
 2IUB_A ~ i i r k k R A D Y L L S L I D A L V D D Y F V L L E K I D D E I D V L E E V I e r p ~
 gi 167593382 ~ r l m n r G P D Y L L Y A I V D G I I D E Y F P I V E R L G D R I D E L D E L y i n p ~
 gi 297551731 ~ q q l h w G T G I L L Y S F L D T L V D G Y F P V V D A L V D Q T E D I E N Q F y i r r g ~
 gi 334090453 ~ q k l e k G L D Y L L Y L V V D S L V Q Y F P L F L K I S D E L E R L E A E Y f a q l ~
 gi 297145379 ~ e f m n r G P D Y L L Y S I I D N I V D D Y F P I E R L A D R I D E L E D D I f t s a ~
 gi 219541761 ~

The highlighted multiple sequence alignment at the end of the record display can be adjusted using the controls at the top (E). The yellow column highlight pattern will be updated to reflect selection in the "Conserved Features/Sites" section (F) at the top.

Reformat | Format: Hypertext | Row Display: up to 10 | Color Bits: 2.0 bit | Type Selection: the most diverse members

Feature 3
 2HN2_A 61 TPTWINITGIhrtDVVRVGEFFgIHLPLVDIlnvhQRPKVEFFe
 gi 257170536 26 GTTWIHAEVt-dAELDAIAETfaIHLPLTVEDLnn-gVRAKTETFF
 gi 353194386 27 GTTWVRVTPs-eAELDEVADGydLHLEIEDVlg-dVRPKVEFFd
 gi 307611750 61 HNIWIDISGLadiEKITKICSEYnIHLPLVIDLIntrQRPKLDIVd
 gi 307156401 65 GSKWVIITGLantQLINDICEFYgMHPLTTEDIlnthQRPKIESF
 gi 257797392 56 WNHIIQIGVhdpAVLEGFRKHfIHLPLADVmntehRPKLEHf
 gi 88187722 66 GICVHISGLkdiDKINQILSDAiLHPLIAEDVfntkGRAKVEDFg
 gi 110620367 27 YNRWISLTRPe-kQIDLVSKFsLHPLVIDItnprEIPKVDEYA
 gi 2649821 63 KKLWIDVGVhdeSLIAKICEFLgIHLPLAEDIIntaQRVKIEDYd
 gi 282157735 27 CNKWVSMSPs-pQEFKAVADAFELHPLVSDMandkELPKVNEYa

Conserved Features/Sites | **PubMed References**

oligomer | **metal binding** | **metal binding** | **CI binding site** | **periplasmic**

Feature 3: metal binding site [ion binding site], 3 residue positions
Conserved feature residue pattern: [ED] [DE] [DE]
Evidence:
 ■ **Structure:** 2IUB: Thermotoga maritime CorA homopentamer binds Mg2+ in metal binding site 2 (M2), contacts at 4A
 - View structure with Cn3D
 ■ **Citation:** PMID 16857941
 ■ **Citation:** PMID 16902408
 ■ **Citation:** PMID 17619822
[Download Cn3D for Viewing 3D Structure](#) | [Scroll to Sequence Alignment Display](#)

Feature 3
 2HN2_A 137 nCVLMFQEKIGDVFDPVREIRynrGIIRKKRADYLLSLIDALVDDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 216
 gi 257170536 104 dWVVSVPAGgAVDVRVRHAvEredERLLQHGADFTAYRVIEVVDDYFVLDREVLEDAEIDEVIDspdreTLERINA 183
 gi 353194386 105 gWLVTEPERLPALDAWVERVgrddPRLLSYGADFTAYRGIDRVDDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 184
 gi 307611750 137 kLLTFRESGYNLSSIIYQRLngehSLTREYGSQYLLFYIMDYVDDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 216
 gi 307156401 140 dTIITFLENDRIEFLIRRLtaskGRLRKNGNDYLFYAIISVVDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 216
 gi 257797392 131 dVLLTFQEGGRDIFLPHARLakpsSRLRQRDLVLYSLDVTVDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 210
 gi 88187722 140 kLLISSETS-NPFENIKTRIsLpnGKFRTHVDVLYMYSIISVVDYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 218
 gi 110620367 100 dFLISASDHW-DTIRTAETGLinkaGIISSEGPDLFAYTLLOHATDHFYVPLDRLEDLIEEIDEAMAtpgkeLLSKMAD 178
 gi 2649821 136 nLVATFEERyWLDISRSRLks-gGRMRKLAGDYLAYTLLOAVVDSYFVLLKIDDEIDVLEEEVLErpekeTVQRTHQ 213
 gi 282157735 99 dFLISLTGNW-DVLRVADATLnskdDPIQKHGIDYALALIDRSVDRFYPIILDDLEDVSDVEERVMGkpakeIIGTISE 177

